REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
13-05-2015	Final Report	26-Jan-2012 - 25-Jan-2015

4. TITLE AND SUBTITLE
Final Report: An Integrated Framework to Access and Mine Distributed Heterogeneous Data Streams with Uncertainty

5a. CONTRACT NUMBER
W911NF-12-1-0066

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
206022

6. AUTHORS
kun zhang

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES
Xavier University of Louisiana
1 Drexel Drive

New Orleans, LA 70125 -1098

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
60449-MA-REP.24

12. DISTRIBUTION AVAILIBILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT
In summary, this DOD agreement not only allowed us to perform a comprehensive study on data stream mining and its promising application on biomedical domains, but also foster and enrich the research experiences of the under-represented minority students at Xavier, and open opportunities for them in graduate schools or future careers in IT industries. Moreover, this support has motivated us to explore other important aspects of mining streaming data, such as anomaly or outlier detection, which is worth more years of further investigation as described in the PI's proposal submitted to the "Fiscal Year 2015 Department of Defense Research and Education Program for

15. SUBJECT TERMS
data streams, data mining, uncertainty, dynamic feature group weighting

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Kun Zhang
UU	UU	UU	UU		19b. TELEPHONE NUMBER: 504-520-6700

Report Title

Final Report: An Integrated Framework to Access and Mine Distributed Heterogeneous Data Streams with Uncertainty

ABSTRACT

In summary, this DOD agreement not only allowed us to perform a comprehensive study on data stream mining and its promising application on biomedical domains, but also foster and enrich the research experiences of the under-represented minority students at Xavier, and open opportunities for them in graduate schools or future careers in IT industries. Moreover, this support has motivated us to explore other important aspects of mining streaming data, such as anomaly or outlier detection, which is worth more years of further investigation as described in the PI's proposal submitted to the "Fiscal Year 2015 Department of Defense Research and Education Program for Historically Black Colleges and Universities and Minority-Serving Institutions (HBCU/MI)". Below are the key statistics achieved by our project.
Peer-reviewed publications: 11 from the PI's lab and 42 from the Co-PI's site
Trained minority undergrads: 20
# of graduates: 11; 4 working in IT industry and 2 attending graduate school
Developed new courses: 2

Enter List of papers submitted or published that acknowledge ARO support from the start of the project to the date of this printing. List the papers, including journal references, in the following categories:

(a) Papers published in peer-reviewed journals (N/A for none)

Received		Paper
05/13/2015	19.00	Wensheng Zhang, Andrea Edwards, Erik Flemington, Kun Zhang, Shannon M. Hawkins. Somatic Mutations Favorable to Patient Survival Are Predominant in Ovarian Carcinomas, PLoS ONE, (11 2014): 1. doi: 10.1371/journal.pone.0112561
08/27/2012	3.00	Wensheng Zhang, Andrea Edwards, Wei Fan, Erik K. Flemington, Kun Zhang. miRNA-mRNA Correlation-Network Modules in Human Prostate Cancer and the Differences between Primary and Metastatic Tumor Subtypes, PLoS ONE, (06 2012): 0. doi: 10.1371/journal.pone.0040130
08/29/2013	5.00	S. Yang, S. Pounds, Kun. Zhang, Z. Fang. PAIR: paired allelic log-intensity-ratio-based normalization method for SNP-CGH arrays, Bioinformatics, (11 2012): 0. doi: 10.1093/bioinformatics/bts683
08/29/2013	6.00	Zhide Fang, Ruofei Du, Andrea Edwards, Erik K. Flemington, Kun Zhang, Yan Gong. The Sequence Structures of Human MicroRNA Molecules and Their Implications, PLoS ONE, (01 2013): 0. doi: 10.1371/journal.pone.0054215
08/30/2014	12.00	Xiaoxiao Shi, Jean-Francois Paiement, David Grangier, Philip S. Yu. GBC: Gradient boosting consensus model for heterogeneous data†, , (06 2014): 0. doi: 10.1002/sam.11193
08/30/2014	13.00	Andrea Edwards, Wei Fan, Wensheng Zhang, Zhide Fang, Prescott Deininger, Kun Zhang. Inferring the expression variability of human transposable element-derived exons by linear model analysis of deep RNA sequencing data, BMC Genomics, (08 2013): 0. doi: 10.1186/1471-2164-14-584
08/30/2014	14.00	Wensheng Zhang, Andrea Edwards, Erik K. Flemington, Kun Zhang, Peter Csermely. Inferring Polymorphism-Induced Regulatory Gene Networks Active in Human Lymphocyte Cell Lines by Weighted Linear Mixed Model Analysis of Multiple RNA-Seq Datasets, PLoS ONE, (10 2013): 0. doi: 10.1371/journal.pone.0078868
08/30/2014	15.00	Bo Liu, Yanshan Xiao, Philip S. Yu, Zhifeng Hao, Longbing Cao. An Efficient Approach for Outlier Detection with Imperfect Data Labels, IEEE Transactions on Knowledge and Data Engineering, (07 2014): 0. doi: 10.1109/TKDE.2013.108
08/30/2014	16.00	Bo Liu, Yanshan Xiao, Philip S. Yu, Longbing Cao, Yun Zhang, Zhifeng Hao. Uncertain One-Class Learning and Concept Summarization Learning on Uncertain Data Streams, IEEE Transactions on Knowledge and Data Engineering, (02 2014): 0. doi: 10.1109/TKDE.2012.235
TOTAL:		**9**

Number of Papers published in peer-reviewed journals:

(b) Papers published in non-peer-reviewed journals (N/A for none)

Received	Paper

TOTAL:

Number of Papers published in non peer-reviewed journals:

(c) Presentations

Number of Presentations: 4.00

Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received	Paper

TOTAL:

Number of Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received		Paper
05/13/2015	18.00	Ke Wu, Kun Zhang, Wei Fan, Andrea Edwards, Philip S. Yu. RS-Forest: A Rapid Density Estimator for Streaming Anomaly Detection, 2014 IEEE International Conference on Data Mining (ICDM). 13-DEC-14, Shenzhen, China. : ,
05/13/2015	23.00	Wensheng Zhang, , Andrea Edwards, , Prescott Deininger, , Kun Zhang. The Duplication and Intragenic Domain Expansion of Human C2H2 Zinc Finger Genes Are Associated with Transposable Elements and Relevant to the Expression-based Clustering, BICoB-2015. 05-MAR-15, . : ,
05/13/2015	22.00	Lifang He, Xiangnan Kong, Philip S. Yu, Zhifeng Hao, Bokai Cao, Ann B. Ragin. Tensor-Based Multi-view Feature Selection with Applications to Brain Diseases, 2014 IEEE International Conference on Data Mining (ICDM). 13-DEC-14, Shenzhen, China. : ,
05/13/2015	21.00	Bokai Cao, Xiangnan Kong, Philip S. Yu. Collective Prediction of Multiple Types of Links in Heterogeneous Information Networks, 2014 IEEE International Conference on Data Mining (ICDM). 13-DEC-14, Shenzhen, China. : ,
05/13/2015	20.00	Jiawei Zhang, Philip S. Yu, Zhi-Hua Zhou. Meta-path based multi-network collective link prediction, the 20th ACM SIGKDD international conference. 23-AUG-14, New York, New York, USA. : ,
08/25/2012	1.00	Sihong Xie, Guan Wang, Shuyang Lin, Philip S. Yu. Review Spam Detection via Temporal Pattern Discovery, the 18th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 12-AUG-12, . : ,
08/25/2012	2.00	Guan Wang, Yuchen Zhao, Xiaoxiao Shi , Philip S. Yu. Magnet Community Identification on Social Networks, the 18th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 12-AUG-12, . : ,
08/29/2013	4.00	jing Peng, Kun Zhang. A Margin Technique for Dimension Reduction with Applications to Hyperspectral Imagery, International Conference on Advanced Computer Science and Electronics Information (ICACSEI 2013). 20-MAY-13, . : ,
08/29/2013	7.00	Yanshan Xiao, Philip S. Yu, Zhifeng Hao, Bo Liu. MODS: Multiple One-class Data Streams Learning from Homogeneous Data, SIAM Data Mining Conference, 2013. 02-MAY-13, . : ,
08/29/2013	8.00	Yuchen Zhao, Philip S. Yu. On Graph Stream Clustering with Side Information, SIAM Data Mining Conference, May 2013. 02-MAY-13, . : ,
08/29/2013	9.00	Bo Liu, Yanshan Xiaoy , Philip S. Yu, Longbing Cao, Zhifeng Hao. Robust Textual Data Streams Mining Based on Continuous Transfer Learning, SIAM Data Mining Conference, 2013. 02-MAY-13, . : ,
08/29/2013	10.00	Chang-Dong Wang, Jian-Huang Lai, Philip S. Yu. Dynamic Community Detection in Weighted Graph Streams, SIAM Data Mining Conference, 2013. 02-MAY-13, . : ,

08/29/2013 11.00 Xiaoxiao Shi└ , Philip Yu. Dimensionality Reduction on Heterogeneous Feature Space, IEEE Intl. Conf. on Data Mining, 2012.. 10-DEC-12, . : ,

08/30/2014 17.00 Ke Wu└ , Andrea Edwards, └Wei Fan, Jing Gao, Kun Zhang└. 2. Classifying Imbalanced Data Streams via Dynamic Feature Group Weighting with Importance Sampling , the 14th SIAM International Conference on Data Mining. 24-APR-14, . : ,

TOTAL: **14**

Number of Peer-Reviewed Conference Proceeding publications (other than abstracts):

(d) Manuscripts

Received Paper

TOTAL:

Number of Manuscripts:

Books

Received Book

TOTAL:

Received Book Chapter

TOTAL:

Patents Submitted

Patents Awarded

Awards

Graduate Students

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Names of Post Doctorates

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Names of Faculty Supported

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Names of Under Graduate students supported

NAME	PERCENT_SUPPORTED	Discipline
Chris chance	0.20	CS
Chris Cosey	0.20	CS
Brittney Mack	0.20	CS
Tuan Nguyen	0.10	CS
Milton Torrey	0.10	CS
Wesley Walker	0.30	CS
FTE Equivalent:	**1.10**	
Total Number:	**6**	

Student Metrics

This section only applies to graduating undergraduates supported by this agreement in this reporting period

The number of undergraduates funded by this agreement who graduated during this period:	11.00
The number of undergraduates funded by this agreement who graduated during this period with a degree in science, mathematics, engineering, or technology fields:	11.00
The number of undergraduates funded by your agreement who graduated during this period and will continue to pursue a graduate or Ph.D. degree in science, mathematics, engineering, or technology fields:	2.00
Number of graduating undergraduates who achieved a 3.5 GPA to 4.0 (4.0 max scale):	1.00
Number of graduating undergraduates funded by a DoD funded Center of Excellence grant for Education, Research and Engineering:	0.00
The number of undergraduates funded by your agreement who graduated during this period and intend to work for the Department of Defense	0.00
The number of undergraduates funded by your agreement who graduated during this period and will receive scholarships or fellowships for further studies in science, mathematics, engineering or technology fields:	2.00

Names of Personnel receiving masters degrees

NAME

Total Number:

Names of personnel receiving PHDs

NAME

Total Number:

Names of other research staff

NAME PERCENT_SUPPORTED

FTE Equivalent:
Total Number:

Sub Contractors (DD882)

Inventions (DD882)

Scientific Progress

Technology Transfer

Final Progress Report (Jan. 2012 – Jan. 2015)

Submitted to the US Army Research Office

Contract #: W911NF-12-1-0066

PI: Kun Zhang, PhD

Computer Science Department

Xavier University of Louisiana

May 13th, 2015

Summary of Project Objective

From January 2012 to January 2015, Dr. Kun Zhang (the PI) at Xavier University of Louisiana and Dr. Philip Yu (the Co-PI) of the University of Illinois at Chicago received grant funding (W911NF-12-1-0066) from the US Army Research Office to develop "An Integrated Framework to Access and Mine Distributed Heterogeneous Data Streams".

The objective of the proposed research is twofold. First, from the scientific aspect, we aimed to develop a distributed, optimization-based, robust data stream mining framework to assist in the battlefield decision-making. By intelligently collecting and analyzing real-time, diversified and uncertain streaming data, multi-source heterogeneous information, as well as constraints as dictated by previous experience and strategic policies, the developed framework is expected to produce the winning strategies as the synthesis of dynamic information gathering and decision-making optimized over a pre-specified loss function, with constraints imposed on interacting sub-components in each single step. Second, from the educational perspective, we intended to offer a vital opportunity to involve African-American students in modern computational research. Educational goals of initiating "data mining" and "mining data streams" classes at the undergraduate level, and recruiting, advising and training undergraduate researchers are also integral to the proposed project.

Project Accomplishments

In order to achieve this objective, the collaborative team conducted a very active research with **53 peer-reviewed publications**. More specifically, the PI's work on this project includes:

- The definition of a margin technique for dimension reduction with applications to hyperspectral imagery [1].
- The design of an importance sampling driven, dynamic feature group weighting framework (DFGW-IS) for classifying data streams of imbalanced distribution [2].
- The development of a novel one-class, semi-supervised algorithm to detect anomalies in streaming data [3].
- By extending the above techniques, the PI also
    - discovered miRNA-mRNA correlation-network modules in human prostate cancer and revealed the differences between primary and metastatic tumor subtypes [4];
    - developed an allelic log-intensity-ratio based normalization method for SNP-CGH arrays [5];
    - explored the sequence structures of human MicroRNA molecules and their implications [6];
    - inferred the expression variability of human transposable element-derived exons by linear model analysis of deep RNA sequencing data[7];
    - constructed polymorphism-induced regulatory gene networks active in human lymphocyte cell lines by weighted linear mixed model analysis of multiple RNA-Seq datasets[8];

- revealed the miRNA-mediated relationships between Cis-SNP genotypes and transcript intensities in lymphocyte cell lines [9];
- identified that somatic mutations favorable to patient survival are predominant in ovarian carcinomas [10];
- investigated the relationship between transposable elements and the duplication and intragenic domain expansion of human C2H2 Zinc finger genes [11].

On the side of Co-PI, the representative work includes “GBC: Gradient Boosting Consensus Model for Heterogeneous Data” [12], “An Efficient Approach for Outlier Detection with Imperfect Data Labels” [13], “Uncertain One-Class Learning and Concept Summarization Learning on Uncertain Data Streams” [14], "Collective Prediction of Multiple Types of Links in Heterogeneous Information Networks" [15], and the other 38 studies [16-53].

From 2012 to 2015, funded by this agreement, **20 Xavier STEM undergraduate researchers** were involved in the proposed project. 95% of those students were African Americans, and 25% of them were female. During this period, we had 11 undergraduates who graduated. Four of those graduates went to work in the IT industry, such as Microsoft and IBM; and two of them are attending graduate schools to pursue a Master or Ph.D in computer science or information system. Please refer to **Appendix 1** for the details of the funded students.

In addition, the PI also introduced **two new courses** into the curriculum. One is the data mining course with the target audience being junior and senior major students, and the other is an interdisciplinary course titled “CPSC2900 Introduction to Bioinformatics Programming”. Designed for students with minimal a priori programming experience, this freshman-level course aims to offering the fundamental bioinformatics programming skills necessary to exploit the abundance of biological data. Both courses have been fully developed and are offered in spring 2015. Sample lectures and problem sets were also given to the STEM students in the summer research seminars with positive feedback.

In summary, this DOD agreement not only allowed us to perform a comprehensive study on data stream mining and its promising application on biomedical domains, but also foster and enrich the research experiences of the under-represented minority students at Xavier, and open opportunities for them in graduate schools or future careers in IT industries. Moreover, this support has motivated us to explore other important aspects of mining streaming data, such as anomaly or outlier detection, which is worth more years of further investigation as described in the PI’s proposal submitted to the “Fiscal Year 2015 Department of Defense Research and Education Program for Historically Black Colleges and Universities and Minority-Serving Institutions (HBCU/MI)”.

Publications

1. J. Peng and **K. Zhang**, "A Margin Technique for Dimension Reduction with Applications to Hyperspectral Imagery", International Conference on Advanced Computer Science and Electronics Information (ICACSEI 2013), May 2013
2. Wu K., Edwards A., Fan W., Gao J. and **Zhang K***.(2014), "Classifying Imbalanced Data Streams via Dynamic Feature Group Weighting with Importance Sampling", Proceedings of the 14th SIAM International Conference on Data Mining. pp. 722-730
3. Wu K., **Zhang K***., Fan W., Edwards A. and Yu P. (2014),"RS-Forest: A Rapid Density Estimator for Streaming Anomaly Detection", Proceedings of the 14th IEEE International Conference on Data Mining
4. W. Zhang, A. Edwards, W. Fan, EK Flemington, **K. Zhang***, "miRNA-mRNA Correlation-Network Modules in Human Prostate Cancer and the Differences between Primary and Metastatic Tumor Subtypes". PLoS ONE 7(6): e40130. doi:10.1371/journal.pone.0040130
5. S. Yang, S. Pounds, **K. Zhang***, and Z. Fang*(2013), "PAIR: Paired Allelic log-Intensity-Ratio based normalization algorithm for SNP-CGH arrays", Bioinformatics. 2013 Feb 1;29(3):299-307. PMID: 23196989, (Journal Impact Factor: 5.323)
6. Fang Z, Du R, Edwards A, Flemington EK, **Zhang K*** (2013) "The Sequence Structures of Human MicroRNA Molecules and Their Implications". PLoS ONE 8(1): e54215. doi:10.1371/journal.pone.005421, PMID: 23349828, (Journal Impact Factor: 4.41)
7. Zhang W., Edwards A., Fan W., Fang Z., Deininger P. and **Zhang K.***(2013), "Inferring the expression variability of human transposable element-derived exons by linear model analysis of deep RNA sequencing data", BMC Genomics, 14:584 doi:10.1186/1471-2164-14-584
8. W. Zhang, A. Edwards, E. Flemington and **K. Zhang***, "Inferring polymorphism-induced regulatory gene networks active in human lymphocyte cell lines by weighted linear mixed model analysis of multiple RNA-Seq datasets", PLOS ONE, 10.1371/journal.pone.0078868
9. W. Zhang, A. Edwards, D. Zhu, EK Flemington, P. Deininger, **K. Zhang***, "miRNA-Mediated Relationships between Cis-SNP Genotypes and Transcript Intensities in Lymphocyte Cell Lines". PLoS ONE 7(2): e31429. doi:10.1371/journal.pone.0031429
10. Zhang W, Edwards A, Flemington E, **Zhang K*** (2014) "Somatic Mutations Favorable to Patient Survival Are Predominant in Ovarian Carcinomas". PLoS ONE 9(11): e112561. doi:10.1371/journal.pone.0112561
11. Zhang W, Edwards A, Flemington E, **Zhang K*** (2015) "The Duplication And Intragenic Domain Expansion of Human C2H2 Zinc Finger Genes Are Associated with Transposable Elements And Relevant to The Expression-based Clustering". BICoB-2015
12. X. Shi, J. Paiement, and D. Grangier, and P.S. Yu, "GBC: Gradient Boosting Consensus Model for Heterogeneous Data", Statistical Analysis and Data Mining, Vol. 7, No. 3, 2014, pp. 161-174.
13. B. Liu, Y. Xiao, P.S. Yu, Z. Hao, and L. Cao "An Efficient Approach for Outlier Detection with Imperfect Data Labels", IEEE Trans. on Knowledge and Data Engineering, Vol. 26, No. 7, July 2014, pp. 1602-1616.
14. B. Liu, Y. Xiao, P.S. Yu, L. Cao, Y. Zhang, and Z. Hao, "Uncertain One-Class Learning and Concept Summarization Learning on Uncertain Data Streams", IEEE Trans. on Knowledge and Data Engineering, Vol. 26, No. 2, Feb. 2014, pp. 468-484.
15. B. Cao, X. Kong, and P.S. Yu, "Collective Prediction of Multiple Types of Links in Heterogeneous Information Networks", Proc. IEEE ICDM Conference, Shenzhen, China, Dec. 2014.
16. B. Cao, L. He, X. Kong, P.S. Yu, Z. Hao, and A. Ragin, "Tensor-based Multi-view Feature Selection with Applications to Brain Diseases", Proc. IEEE ICDM Conference, Shenzhen, China, Dec. 2014.
17. S. Lin, Q. Hu, F. Wang, and P.S. Yu, "Steering Information Diffusion Dynamically against User Attention Limitation", Proc. IEEE ICDM Conference, Shenzhen, China, Dec. 2014.

18. L. He, X. Yang, H. Shuai, X. Kong, and P.S. Yu, "Low-Density Cut based Tree Decomposition for Large-Scale SVM Problems", Proc. IEEE ICDM Conference, Shenzhen, China, Dec. 2014.
19. J. Zhang, X. Kong, R. Luo, Y. Chang, and P.S. Yu, "NCR: A Scalable Network-Based Approach to Co-Ranking in Question-and-Answer Sites", Proc. ACM CIKM Conference, Shanghai, China, Oct. 2014.
20. C. Lu, H. Shuai, and P.S. Yu, "Identifying Your Customers in Social Networks", Proc. ACM CIKM Conference, Shanghai, China, Oct. 2014.
21. J. Zhang, P.S. Yu, and Z. Zhou, "Meta-path based Multi-network Collective Link Prediction", Proc. ACM KDD Conference, New York, NY, Aug. 2014.
22. B. Liu, Y. Xiao, P.S. Yu, Y. Zhang, and L. Cao, "An Efficient Orientation Distance-Based Discriminative Feature Extraction Method for Multi-Class Classification Problems", Knowledge and Information Systems Journal, Vol. 39, No. 2, May 2014, pp. 409-433.
23. C. Tai, P. Tseng, P.S. Yu, and M.S. Chen, "Identity Protection in Sequential Releases of Dynamic Social Networks", IEEE Trans. on Knowledge and Data Engineering, Vol. 26, No, 3, March 2014, pp. 635-651.
24. C. Aggarwal and Y. Xie, and P.S. Yu, "A Framework for Dynamic Link Prediction in Heterogeneous Networks", Statistical Analysis and Data Mining, Vol. 7, No. 1, Feb. 2014, pp. 14-33.
25. C. Tai, P.S. Yu, D. Yang, and M.S. Chen "Structural Diversity for Resisting Community Identification in Published Social Networks", IEEE Trans. on Knowledge and Data Engineering, Vol. 26, No. 1, Jan. 2014, pp. 235-252.
26. H. Shuai, D. Yang, P.S. Yu, and M.S. Chen "Willingness Optimization for Social Group Activity", PVLDB, Vol. 7, No. 4, Dec., 2013.
27. C. Wang, J. Lai, and P.S. Yu, "NEIWalk: Community Discovery in Dynamic Content-based Networks", IEEE Trans. on Knowledge and Data Engineering, Vol. 26, No. 7, July 2014, pp. 1734-1748
28. R. Chen, B. Fung, P.S. Yu, and B. Desai "Correlated Network Data Publication via Differential Privacy", VLDB Journal, Vol. 23, No. 4, 2014, pp. 635-676.
29. C. Aggarwal, Y. Zhao, and P.S. Yu, "On the use of Side Information for Mining Text Data", (with), IEEE Trans. on Knowledge and Data Engineering, Vol. 26, No. 6, June 2014, pp. 1415-1429.
30. J. Zhang, and P.S. Yu, "Meta-path based Multi-network Collective Link Prediction", Proc. ACM KDD Conference, New York, NY, Aug. 2014.
31. L. He, X. Kong, P.S. Yu, A. Ragin, Z. Hao, and X. Yang, "DuSK: A Dual Structure-preserving Kernel for Supervised Tensor Learning with Applications to Neuroimages", Proc. SIAM Data Mining Conference, Philadelphia, PA, April 2014.
32. N. Yang, X. Kong, F. Wang, and P.S. Yu, "When and Where: Predicting Human Movements Based on Social Spatial-Temporal Events", Proc. SIAM Data Mining Conference, Philadelphia, PA, April 2014.
33. P. Peng, R. Wong, and P.S. Yu, "Learning on Probabilistic Labels", Proc. SIAM Data Mining Conference, Philadelphia, PA, April 2014.
34. S. Wang, S. Xie, Z. Li, and P.S. Yu, "Future Influence Ranking of Scientific Literature", Proc. SIAM Data Mining Conference, Philadelphia, PA, April 2014.
35. X. Kong, Z. Wu, L. Li, R. Zhao, H. Wu, W. Fan, and P.S. Yu, "Large-Scale Multi-Label Learning with Incomplete Label Assignments", Proc. SIAM Data Mining Conference, Philadelphia, PA, April 2014.
36. C. Lu, S. Xie, X. Kong, and P.S. Yu, "Inferring the Impacts of Social Media on Crowdfunding", Proc. ACM WSDM Conference, New York, NY, Feb. 2014.
37. J. Zhang, X. Kong, and P.S. Yu, "Transferring Heterogeneous Links across Location-Based Social Networks", Proc. ACM WSDM Conference, New York, NY, Feb. 2014.
38. X. Zhu, S. Song, J. Wang, J. Sun, and P.S. Yu, "Matching Heterogeneous Events with Patterns", Proc. IEEE Intl. Conf. on Data Engineering, Chicago, IL, April 2014.

39. H. Shuai, D. Yang, P.S. Yu, C. Shen, and M.S. Chen, "On Pattern Preserving Graph Generation", Proc. IEEE Intl. Conf. on Data Mining, Dallas, TX, Dec. 2013.
40. S. Xie, X. Kong, J. Gao, W. Fan, and P.S. Yu, "Multilabel Consensus Classification", Proc. IEEE Intl. Conf. on Data Mining, Dallas, TX, Dec. 2013.
41. W. Shao, X. Shi, and P.S. Yu, "Clustering on Multiple Incomplete Datasets via Collective Kernel Learning", Proc. IEEE Intl. Conf. on Data Mining, Dallas, TX, Dec. 2013.
42. J. Zhang, X. Kong, and P.S. Yu, "Predicting Social Links for New Users across Aligned Heterogeneous Social Networks", Proc. IEEE Intl. Conf. on Data Mining, Dallas, TX, Dec. 2013.
43. M. Long, J. Wang, G. Ding, and P.S. Yu, "Transfer Feature Learning with Joint Distribution Adaptation", Proc. Intl. Conference on Computer (ICCV), Sydney, Australia, Dec. 2013.
44. X. Kong, J. Zhang, and P.S. Yu, "Inferring Anchor Links across Multiple Heterogeneous Social Networks", Proc. ACM CIKM Conference, San Francisco, Oct. 2013.
45. S. Lin, X. Kong, and P.S. Yu, "Predicting Trends in Social Networks via Dynamic Activeness Model", Proc. ACM CIKM Conference, San Francisco, Oct. 2013.
46. Y. Zhao, and P.S. Yu, "On Graph Stream Clustering with Side Information", Proc. SIAM Data Mining Conference, Austin, TX, May 2013.
47. C. Wang, J. Lai, and P.S. Yu, "Dynamic Community Detection in Weighted Graph Streams", Proc. SIAM Data Mining Conference, Austin, TX, May 2013.
48. B. Liu, Y. Xiao, and P.S. Yu, "MODS: Multiple One-class Data Streams Learning from Homogeneous Data", Proc. SIAM Data Mining Conference, Austin, TX, May 2013.
49. B. Liu, Y. Xiao, P.S. Yu, and L. Cao, "Robust Textual Data Streams Mining Based on Continuous Transfer Learning", Proc. SIAM Data Mining Conference, Austin, TX, May 2013.
50. B. Liu, Y. Xiao, and P.S. Yu, "One-Class Transfer Learning with Uncertain Data", Proc. Pacific-Asia Conf. on Knowledge Discovery and Data Mining (PAKDD), Gold Coast, Australia, April 2013.
51. X. Shi, and P.S. Yu, "Dimensionality Reduction on Heterogeneous Feature Space", Proc. IEEE Intl. Conf. on Data Mining, Brussels, Belgium, Dec. 2012.
52. S. Xie, G. Wang, S. Lin, and P.S. Yu, "Review Spam Detection via Temporal Pattern Discovery", Proc. ACM KDD Conference, Beijing, China, Aug. 2012.
53. G. Wang, Y. Zhao, and X. Shi, and P.S. Yu, "Magnet Community Identification on Social Networks", Proc. ACM KDD Conference, Beijing, China, Aug. 2012.